

WHAT IS CLAIMED IS:

1 1. A system for retrieving and integrating data from a plurality of data
2 sources, comprising:
3 an aggregation server configured to convert a data request into an internal
4 query and generate one or more subqueries from the internal query by matching the internal
5 query against a set of rules; and
6 one or more agents, each agent configured to communicate with one or more
7 data sources and retrieve data from the one or more data sources pursuant to an associated
8 subquery provided by the aggregation server;
9 wherein the aggregation server is further configured to join, fuse and union
10 respective data retrieved by the one or more agents; and
11 wherein the one or more agents are located at respective remote locations and
12 the aggregation server communicates with the one or more agents via a computer network.

1 2. The system of claim 1 wherein:
2 the internal query is represented by a query definition file having a head
3 portion and a tail portion;
4 the set of rules is represented by a mediator specification file, each rule within
5 the mediator specification file having a head portion and specifying how one of a plurality of
6 internal queries is satisfied by one or more of the plurality of data sources;
7 a subset of rules within the set of rules is considered to be a match with the
8 internal query if the tail portion of the query definition file matches the combined set of head
9 portions of the subset of rules; and
10 for each set of matched rules, the aggregation server generates a
11 corresponding subquery.

1 3. The system of claim 1 wherein each agent has a corresponding agent
2 capability file; and
3 wherein the aggregation server is further configured to check the
4 corresponding agent capability file of an agent before the agent is invoked to retrieve data
5 from one or more data sources pursuant to an associated subquery provided by the
6 aggregation server.

1 4. The system of claim 1 wherein upon receiving the associated subquery
2 provided by the aggregation server, an agent uses a query mapping file which corresponds to
3 the associated subquery to enable the one or more data sources to be accessed.

1 5. The system of claim 1 wherein the aggregation server formulates a
2 query execution plan using the one or more subqueries generated from the the internal query;
3 and

4 wherein the one or more subqueries are executed by their respective agents
5 pursuant to the query execution plan to optimize access to the one or more data sources.

1 6. The system of claim 1 wherein the computer network is the Internet.

2 7. The system of claim 1 wherein the plurality of data sources includes a
3 database or an application.

4 8. The system of claim 1 wherein the aggregation server and the one or
5 more agents are implemented using software, hardware or a combination of both.

6 9. The system of claim 1 wherein communications between the
7 aggregation server and the one or more agents are encoded in XML format.

8 10. A system for retrieving and integrating data from a plurality of data
9 sources over a computer network, comprising:

10 an aggregation server configured to receive a data request from a user and
11 convert the data request into an internal query, the aggregation server further configured to
12 match the internal query against a set of rules and, for each matched rule, generate a
13 corresponding subquery; and

1 a plurality of agents, one or more agents forming a set of agents configured to
2 communicate with one or more data sources and retrieve data from the one or more data
3 sources pursuant to a corresponding subquery provided by the aggregation server;

4 wherein the aggregation server is further configured to union respective data
5 retrieved by one or more of the plurality of agents; and

6 wherein the plurality of agents are located at respective remote locations and
7 the aggregation server communicates with the plurality of agents via the computer network.

8 11. The system of claim 10 wherein:

the internal query is represented by a query definition file having a head portion and a tail portion;

the set of rules is represented by a mediator specification file, each rule within the mediator specification file having a head portion and specifying how one of a plurality of internal queries is satisfied by one or more of the plurality of data sources; and

a subset of rules within the set of rules is considered to be a match with the internal query if the tail portion of the query definition file matches the combined head portions of the subset of rules.

12. The system of claim 10 wherein each agent has a corresponding agent capability file; and

wherein the aggregation server is further configured to check the corresponding agent capability file of an agent before the agent is invoked to retrieve data from one or more data sources pursuant to the corresponding subquery provided by the aggregation server.

13. The system of claim 10 wherein upon receiving the corresponding subquery provided by the aggregation server, an agent uses a query mapping file which maps the corresponding subquery to enable the one or more data sources to be accessed.

14. The system of claim 10 wherein the aggregation server formulates a query execution plan using the one or more subqueries generated from the the internal query; and

wherein the one or more subqueries are executed by their respective sets of agents pursuant to the query execution plan to optimize access to the one or more data sources.

15. The system of claim 14 wherein the aggregation server formulates the query execution plan by identifying a common data source to be accessed pursuant to the one or more subqueries and a common key shared by respective data sources to be accessed pursuant to the one or more subqueries.

16. The system of claim 10 wherein the computer network is the Internet.

17. The system of claim 10 wherein the plurality of data sources includes a database or an application.

1 18. The system of claim 10 wherein the aggregation server and the
2 plurality of agents are implemented using software, hardware or a combination of both.

1 19. The system of claim 10 wherein communications between the
2 aggregation server and the plurality of agents are encoded in XML format.

1 20. A system for retrieving and integrating data via a computer network,
2 comprising:

3 an aggregation server configured to include a query definition file generated
4 from a data request, the query definition file having a head portion and a tail portion, a
5 mediator specification file having a plurality of rules, each rule having a head portion, and a
6 plurality of agent capability files;

7 a plurality of data sources; and

8 a plurality of agents, each agent configured to retrieve data from one or more
9 of the plurality of data sources and having a corresponding query mapping file;

10 wherein:

11 the aggregation server maintains an agent capability file for each agent;

12 the aggregation server matches the query definition file against the plurality of
13 rules;

14 a set of rules is considered to be a match if the tail portion of the query
15 definition file matches the corresponding head portion of the rule;

16 for each matched rule, the aggregation server generates a corresponding
17 subquery, the subquery including information on a set of specific agents to be invoked;

18 for the specific agents to be invoked, the aggregation server checks the
19 corresponding agent capability files to determine if the specific agents are capable of
20 handling the corresponding subqueries;

21 for specific agents that are determined to be capable of handling their
22 corresponding subqueries, each such specific agent is caused to retrieve data from one or
23 more of the plurality of data sources using its corresponding query mapping file; and

24 upon receiving the retrieved data from the specific agents, the aggregation
25 server performs join, fusion and union operations on the retrieved data.

1 21. The system of claim 20 wherein the aggregation server formulates a
2 query execution plan using the corresponding subqueries; and

wherein the corresponding subqueries are executed by their respective sets of specific agents pursuant to the query execution plan to optimize access to the one or more data sources.

22. The system of claim 21 wherein the aggregation server formulates the query execution plan by identifying a common data source to be accessed pursuant to the corresponding subqueries and a common key shared by respective data sources to be accessed pursuant to the corresponding subqueries.

23. The system of claim 20 wherein the computer network is the Internet.

24. The system of claim 20 wherein the plurality of data sources includes a database or an application.

25. The system of claim 20 wherein the aggregation server and the plurality of agents are implemented using software, hardware or a combination of both.

26. The system of claim 20 wherein communications between the aggregation server and the plurality of agents are encoded in XML format.

27. A method for using an aggregation server and a plurality of agents to retrieve and integrate data from a plurality of data sources via a computer network, comprising:

configuring the aggregation server to perform the following:

receiving a data request from a user;

converting the data requesting into an internal query;

matching the internal query against a plurality of rules;

for each set of matched rules, generating a corresponding subquery for an agent;

forwarding information relating to all generated subqueries to their respective sets of agents;

performing join, fusion and union operations on data returned from the respective sets of agents; and

configuring each of the plurality of agents to perform the following:

receiving information to the corresponding subquery from the aggregation server;

retrieving data pursuant to the corresponding subquery from one or more of the plurality of data sources; and
returning the retrieved data to the aggregation server.

28. The method of claim 27 wherein the step of configuring the aggregation server further comprises:
formulating a query execution plan using all the generated subqueries; and
forwarding information relating to all the generated subqueries to their respective sets of agents pursuant to the query execution plan.

29. The method of claim 28 wherein the step of formulating the query execution plan further comprises:
identifying a common data source to be accessed pursuant to the generated subqueries and a common key shared by respective data sources to be accessed pursuant to the generated subqueries.

30. The method of claim 27 wherein the step of configuring the aggregation server further comprises:
before forwarding information relating to all generated subqueries to their respective sets of agents, checking each respective agent to determine if such agent is capable of handling the corresponding subquery.

31. The method of claim 27 wherein the computer network is the Internet.

32. The method of claim 27 wherein the plurality of data sources includes a database or an application.

33. The method of claim 27 wherein the method is implemented using software, hardware or a combination of both.

34. A method for using an aggregation server and a plurality of agents to retrieve and integrate data from a plurality of data sources via a computer network, comprising:
instructing the aggregation server to generate a query definition file from a data request, the query definition file having a head portion and a tail portion;
instructing the aggregation server to match the query definition file against a mediator specification file having a plurality of rules, each rule having a head portion, a set of rules is considered to be a match if the tail portion of the query definition file matches the corresponding combined head portions of the rules in the set;

10 for each set of matched rules, instructing the aggregation server to generate a
11 corresponding subquery, the subquery including information on a specific set of agents to be
12 invoked;

13 for the specific set of agents to be invoked, instructing the aggregation server
14 to check corresponding agent capability files to determine if each of the specific set of agents
15 is capable of handling the corresponding subqueries;

16 for specific agents that are determined to be capable of handling their
17 corresponding subqueries, instructing each such specific agent to retrieve data from one or
18 more of the plurality of data sources using a corresponding query mapping file and return the
19 retrieve data to the aggregation server; and

20 upon receiving the returned data from the specific agents, instructing the
21 aggregation server to perform join, fusion and union operations on the returned data.

22 35. The method of claim 34 further comprising:

23 instructing the aggregation server to formulate a query execution plan using
24 the corresponding subqueries; and

25 instructing the aggregation server to cause the corresponding subqueries to be
26 executed by their respective specific sets of agents pursuant to the query execution plan to
27 optimize access to the one or more data sources.

28 36. The method of claim 35 wherein the step of instructing the aggregation
29 server to formulate the query execution plan further comprises:

30 identifying a common data source to be accessed pursuant to the
31 corresponding subqueries and a common key shared by respective data sources to be
32 accessed pursuant to the corresponding subqueries.